Applications of Microsoft® Excel in Analytical Chemistry

Second Edition

Stanley R. Crouch
Michigan State University

F. James Holler
University of Kentucky

BROOKS/COLE
CENGAGE Learning

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

Chapter 4

Least-Squares and Calibration Methods

Many of the standardization and calibration methods in analytical chemistry are based on the assumption of a linear relationship between the measured quantity and the concentration of the analyte. In particular, the **external standard method** involves preparing a **calibration curve** or **working curve** from the responses of a set of standards containing the analyte at known concentrations. The parameters of the linear relationship, such as the slope and intercept, are then calculated by least-squares analysis, which is also widely used in other calibration procedures such as the internal standard method and the method of standard additions. This chapter describes the use of Excel for obtaining least-squares estimates of slopes and intercepts and for predicting unknown concentrations. Multiple linear regression and polynomial linear regression are also discussed. Nonlinear regression methods are described in Chapter 13.

Linear Least-Squares Analysis

Linear least-squares analysis is fairly straightforward with Excel. This type of analysis can be accomplished in several ways: by entering the equations for linear regression manually, by employing the built-in functions of Excel, or by utilizing the regression data analysis tool. Because the built-in functions are the easiest of these options, we'll explore them in detail and see how they may be used to evaluate analytical data. We'll also briefly describe the regression analysis tool.

The Slope and Intercept

As an example, we will take the external standard calibration data for the chromatographic determination of isooctane as described in example 7-7 of AC7 or Example 8-4 of FAC9. We enter the data from Table 7-4 of AC7 or 8-1 of FAC9 so that it appears as shown in Figure 4-1.

	A	B	C
1		*x*	*y*
2		0.352	1.09
3		0.803	1.78
4		1.08	2.6
5		1.38	3.03
6		1.75	4.01
7			
8	**Slope**		
9	**Intercept**		
10			

Figure 4-1 Worksheet after entering data.

Now, click on cell B8, and then on the Home tab. Click on the Insert Function icon so that the window shown in Figure 4-2 appears. Then click on the Statistical category.

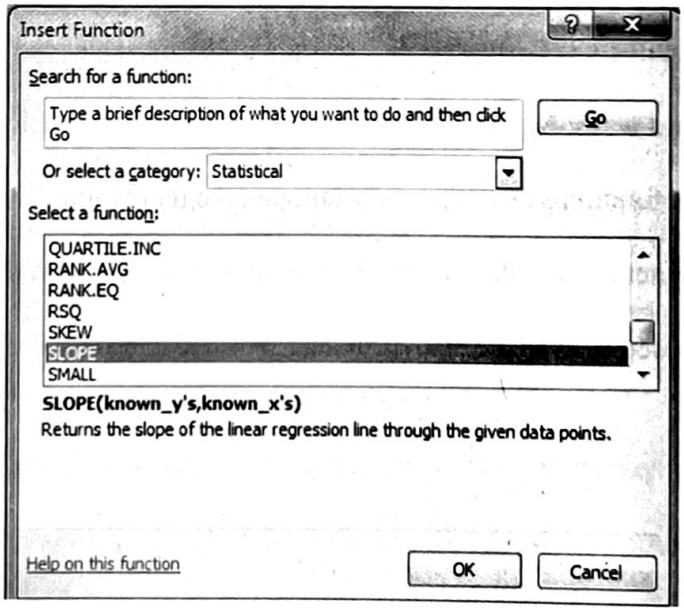

Figure 4-2 Insert Function window.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Use the mouse to scroll down the list of functions until you come to the SLOPE function, and the click on it. The function appears in bold under the Select a function: window, and a description of the function appears below it. Read the description of the slope function, and then click OK. The window shown in Figure 4-3 appears just below the formula bar.

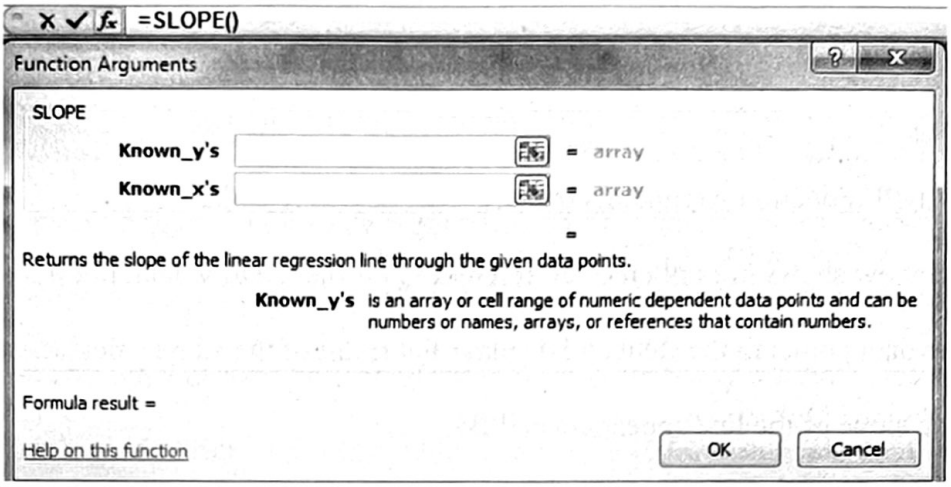

Figure 4-3 Function Arguments window for SLOPE function.

Look carefully at the information that is provided in the window and in the formula bar. The SLOPE() function appears in the formula bar with no arguments, so we must select the data that Excel will use to determine the slope of the line. Now click on the selection button at the right end of the Known_y's box, use the mouse to select cells C2:C6, and type [↵]. Similarly, click on the selection button for the Known_x's box, select cells B2:B6 followed by [↵]. The window should now appear as shown in Figure 4-4.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

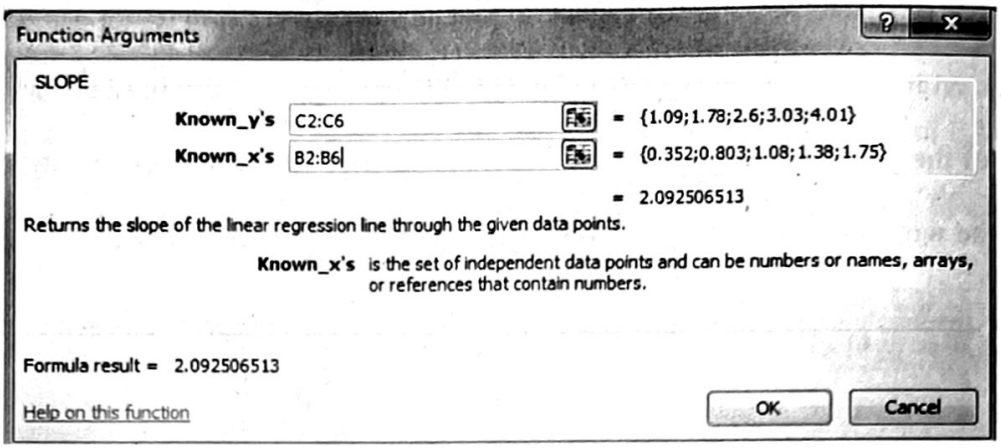

Figure 4-4 SLOPE Function Arguments window after selecting cells.

The window shows not only the cell references for the x and y data, but it also shows the first few of the data points to the right and displays the result of the slope calculation. Now click on OK, and the slope of the line appears in cell B8.

Click on cell B9 followed by the **Insert Function** icon, and repeat the process that we just carried out, except that now you should select the INTERCEPT function. When the intercept Function Arguments window appears, select the Known_y's and the Known_x's as before, and click OK. When you have finished, the worksheet should appear as shown in Figure 4-5.

	A	B	C
1		*x*	*y*
2		0.352	1.09
3		0.803	1.78
4		1.08	2.6
5		1.38	3.03
6		1.75	4.01
7			
8	**Slope**	2.092507	
9	**Intercept**	0.256741	

Figure 4-5 Worksheet after calculation of slope and intercept.

At this point you may wish to compare these results to those obtained for the slope and intercept in Example 8-4 of FAC9 or Example 7-7 of AC7. We should note at this point that Excel

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

provides many digits that are not significant. We shall see how many figures are significant after we find the standard deviations of the slope and intercept.

Using LINEST

Another way to accomplish the least-squares analysis is by using the LINEST function. On the worksheet shown in Figure 4-5, use the mouse to select an array of cells two cells wide and five cells high, such as E2:F6. Then click on the Insert Function icon, select STATISTICAL and LINEST in the left and right windows, respectively, and click on OK. Select the Known_y's and Known_x's as before, then click on the box labeled Const and type **`true`**. Also type **`true`** in the box labeled Stats. When you click on each of the latter two boxes, notice that a description of the meaning of these logical variables appears below the box. In order to activate the LINEST function, you must now type the rather unusual keystroke combination **`Ctrl+Shift+[↵]`**. This keystroke combination must be used whenever you perform a function on an array of cells. The worksheet should now appear as in Figure 4-6.

	A	B	C	D	E	F	G
1		*x*	*y*				
2		0.352	1.09		2.092507	0.256741	
3		0.803	1.78		0.134749	0.158318	
4		1.08	2.6		0.987712	0.144211	
5		1.38	3.03		241.1465	3	
6		1.75	4.01		5.015089	0.062391	
7							
8	**Slope**	2.092507					
9	**Intercept**	0.256741					

Figure 4-6 Worksheet after implementing the LINEST function.

As you can see, cells E2 and F2 contain the slope and intercept of the least-squares line. Cells E3 and F3 are the respective standard deviations of the slope and intercept. Cell E4 contains the coefficient of determination (R^2). The standard deviation about regression (s_r, standard error of

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

the estimate) is located in cell F4. The smaller the s_r value, the better the fit. The square of the standard error of the estimate is the mean square for the residuals (error). The value in cell E5 is the F statistic. Cell F5 contains the number of degrees of freedom associated with the error. Finally cells E6 and F6 contain the sum of the squares of the regression and the sum of the squares of the residuals, respectively. Note that the F value can be calculated from these latter quantities as described in Section 8D-2 of FAC9.

It is worth noting that the number of significant figures that we keep in a least-squares analysis depends on the use for which the data are intended. If the results are to be used to carry out further spreadsheet computations, wait until final results are computed before rounding to an appropriate number of significant figures. Excel provides 15 digits of numerical precision, and so, in general, spreadsheet computations will not contribute to the uncertainty in the final results. Final answers must be rounded to be consistent with the uncertainty in the original data, which is reflected in the standard deviations of the slope and intercept and the standard error of the estimate. The standard deviations of the slope and intercept in our example suggest that, at most, we should express both the slope and the intercept to only two decimal places. Thus, the least-squares results for the slope and intercept may be expressed as 2.09 ± 0.13 and 0.26 ± 0.16, respectively, or as 2.1 ± 0.1 and 0.3 ± 0.2.

The Analysis ToolPak Regression Tool

Yet a third way to do a linear least-squares analysis is to use the regression function in Excel's Analysis ToolPak. The advantage of this third mode is the production of a complete ANOVA table for the analysis. Select the **Data** tab and bring up the Data ribbon. At the far right, select **Data Analysis**. From the Data Analysis window, select **Regression** and then click on OK. Using

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

the isooctane worksheet as before, select C2:C6 as the Input Y Range: and B2:B6 as the Input X Range:. Select New Worksheet Ply: for the output. The results should appear as shown in Figure 4-7. Note you will have to expand some cell widths to see all the text.

In addition to the normal regression statistics, the output of Figure 4-7 shows a statistic called the multiple correlation coefficient (multiple R) and the adjusted R square. The former is beyond the scope of this discussion[1], while the latter is the R^2 value adjusted for the number of parameters used in the fit (the adjusted R^2 includes a "price" to pay for adding more parameters to improve the fit). In the ANOVA table the entries are the degrees of freedom (*df*), the sum of squares (*SS*), the mean square values (*MS*), the *F* statistic, and the significance level of the *F* value (probability of getting an *F* value this large by random chance alone). The bottom table gives the slope and intercept, their standard deviations (standard errors), the *t* statistics for the slope and intercept, the probabilities of getting these *t* values by random chance, and the upper and lower values for the 95% confidence intervals for the slope and intercept. From this analysis, we note that the linear model fits the data quite well. There is, however, significant uncertainty in the intercept value.

[1] For further discussion of the multiple correlation coefficient, see J. L. Devore, *Probability and Statistics for Engineering and the Sciences,* 8th edition (Boston, MA: Duxbury, Brooks/Cole, 2012), p. 560.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	A	B	C	D	E	F	G
1	SUMMARY OUTPUT						
2							
3	*Regression Statistics*						
4	Multiple R	0.993837158					
5	R Square	0.987712297					
6	Adjusted R Square	0.983616396					
7	Standard Error	0.144211147					
8	Observations	5					
9							
10	ANOVA						
11		*df*	*SS*	*MS*	*F*	*Significance F*	
12	Regression	1	5.015089435	5.015089	241.1465	0.000580234	
13	Residual	3	0.062390565	0.020797			
14	Total	4	5.07748				
15							
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
17	Intercept	0.256740511	0.158317598	1.62168	0.203322	-0.247096745	0.760578
18	X Variable 1	2.092506513	0.134749235	15.52889	0.00058	1.663674308	2.521339

Figure 4-7 Results using Analysis ToolPak Regression Tool.

Plotting a Graph of the Data and the Least-Squares Fit

It is customary and useful to plot a graph of the data and the least-squares fitted line. The built-in charting engine of Excel makes creating such plots relatively easy. There are several ways to display the data points and the predicted line simultaneously. One way is to plot the predicted values $\hat{y}$ and the experimental y values simultaneously. The predicted values for the isooctane data are given in Table 8-2 of FAC9. The easiest way is to have Excel add the line, called a trendline, itself.

To plot the points, select the xy data (Cells B2:C6) from the original isooctane worksheet. Click on the **Insert** tab to display the Insert ribbon. The Charts group on the Insert ribbon includes drop-down menus for seven charting types: column, line, pie, bar, area, scatter, and other charts. Since we want to plot xy data, display the **Scatter** drop-down shown in Figure 4-8 by clicking on the downward pointing arrow under the Scatter chart type.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

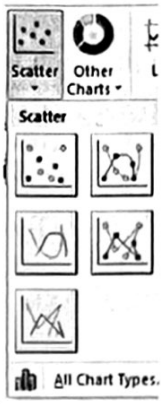

Figure 4-8 Drop-down menu for Scatter chart type.

Choose the upper left Scatter type (Scatter with only markers). This will produce the default graph shown in Figure 4-9.

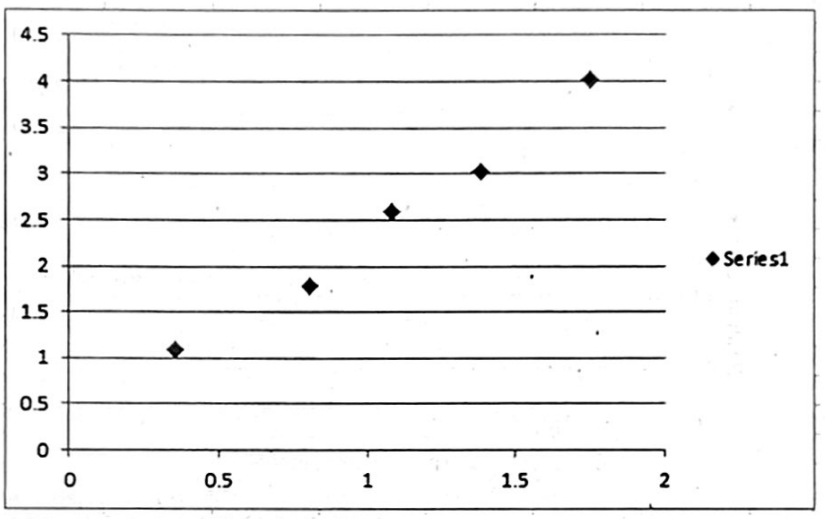

Figure 4-9 Graph of isooctane calibration data.

Next we will make the graph look more presentable for scientific data. Click on the chart and note that the Chart Tools ribbon appears. We will first change the gridlines. Click on the **Layout** tab on the Chart Tools ribbon to display the Layout gallery shown in Figure 4-10.

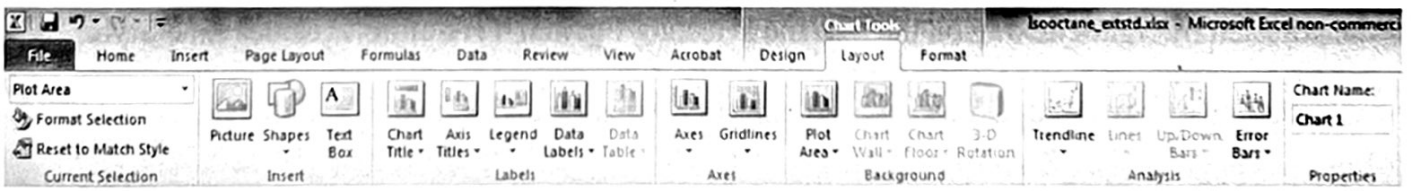

Figure 4-10 Chart layout tab.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Now click on **Gridlines** in the axes group and choose Major gridlines under the Primary Vertical gridlines option. Note that the default graph has already chosen Major gridlines for the Primary Horizontal gridlines. Now click on **Axis Titles** in the Labels group. Under the Primary Horizontal Axis Title, choose Title Below Axis. Type Concentration of isooctane, mol % for the horizontal axis title. Under the Primary Vertical Axis Title, choose **Rotated Title** and type Peak area, arbitrary units. Note that the default font size is 10 point. You can change the font face and font size by selecting the text to be changed, clicking on the Home tab to display the Home ribbon and choosing a new font face and font size in the Font group. Change the Axis Titles to be Arial font face and 11 point font size. Your chart should now appear as shown in Figure 4-11.

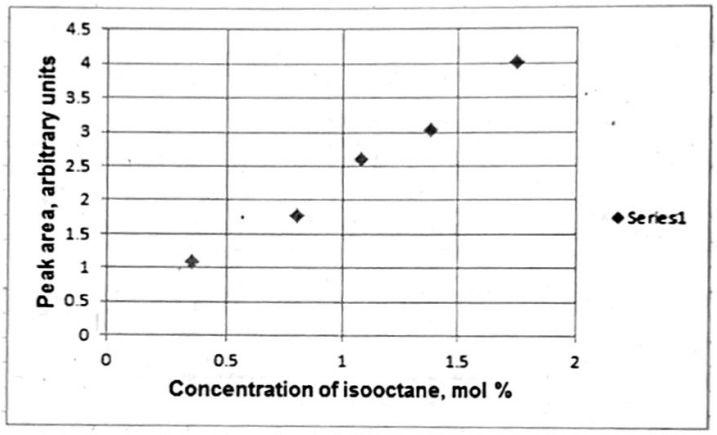

Figure 4-11 Plot of isooctane calibration data with gridlines and axes titles.

Next we will add the least-squares line. Now right-click on any data point and then click on **Add Trendline....** Under Trendline Options, select **Linear**, and check both Display Equation on chart and Display R-squared value on chart. Then click on Close. The weight of the line can be adjusted by right-clicking on the line and selecting **Format Trendline....**Under Line Style, select a line of 1 pt. Width. You can also move the equation and R^2 text to a more convenient place as indicated in Figure 4-12.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

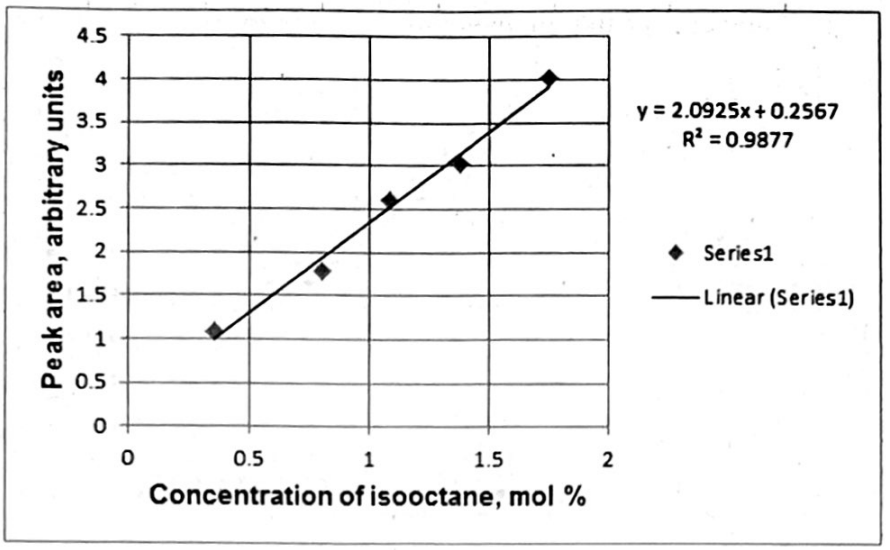

Figure 4-12 Plot of isooctane calibration curve with added trendline.

As an extension of this exercise, modify your spreadsheet to include a column of residuals. You can either enter them from Table 8-2 of FAC9 or include residuals when obtaining the ANOVA table in the Analysis ToolPak. You can even produce the chart of the fit and the residuals directly from the Regression window by checking the appropriate boxes. Be sure to save your spreadsheet in a file for reference and for future use.

Using Excel to Find Unknown Concentrations From a Calibration Curve

Now let's use the calibration curve from above to find the concentration of isooctane in a sample and the standard deviation in concentration. We'll use the data of Example 8-5 of FAC9 or Example 7-8 of AC7. In this example, a peak area of 2.65 was obtained for the unknown. Type the label **Unknown** in cell A7 of the original worksheet, and enter 2.65 in cell C7.

From the equation for a line, $y = mx + b$, we can rearrange to give the concentration x as

$$x = \frac{y-b}{m}$$

Add the label **Concentration of unknown** in cell A10 and enter in cell B10

```
=(C7-B9)/B8[↵]
```


© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

Your worksheet should now appear like that shown in Figure 4-13.

	A	B	C	D	E	F
1		*x*	*y*			
2		0.352	1.09		2.092507	0.256741
3		0.803	1.78		0.134749	0.158318
4		1.08	2.6		0.987712	0.144211
5		1.38	3.03		241.1465	3
6		1.75	4.01		5.015089	0.062391
7	**Unknown**		2.65			
8	**Slope**	2.092507				
9	**Intercept**	0.256741				
10	**Concentration of unknown**	1.143729				

Figure 4-13 Least-squares worksheet after entering unknown and calculating its concentration.

In order to find the standard deviation in concentration, we utilize Equation 8-18 of FAC9 or 7-18 of AC7. In this equation we need several additional quantities. First, we need s_r, the standard deviation about regression, also called the standard error in *Y*. We also need *M*, the number of replicate analyses of unknowns (1 in this case) and *N*, the number of points in the calibration curve (5 in this case). Finally we need S_{xx}, the sum of the squares of the deviations of *x* values from the mean *x* value and the mean *y* value. In column A, add the labels **`standard error in Y`**, ***N***, S_{xx}, **`y bar`**, ***M***, and **`Standard deviation in c`** in cells A11 through A16. Your worksheet should now appear as shown in Figure 4-14.

	A	B	C	D	E	F
1		*x*	*y*			
2		0.352	1.09		2.092507	0.256741
3		0.803	1.78		0.134749	0.158318
4		1.08	2.6		0.987712	0.144211
5		1.38	3.03		241.1465	3
6		1.75	4.01		5.015089	0.062391
7	**Unknown**		2.65			
8	**Slope**	2.092507				
9	**Intercept**	0.256741				
10	**Conc. Unknown**	1.143729				
11	**Standard error in *Y***					
12	***N***					
13	**S_{xx}**					
14	***y* bar**					
15	***M***					
16	**Standard deviation in *c***					

Figure 4-14 Worksheet after entering labels for error analysis.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

The standard error in *Y* could be found from Equation 8-15 of FAC9 or 7-15 of AC7. However, it is easier to make use of the built-in Excel function STEYX(), which returns the standard error in *Y* directly. Hence, select cell B11. Click on the Insert Function icon. In the Insert Function window, select the Statistical category and the STEYX function. Click OK. This opens the STEYX Function Arguments window shown in Figure 4-15.

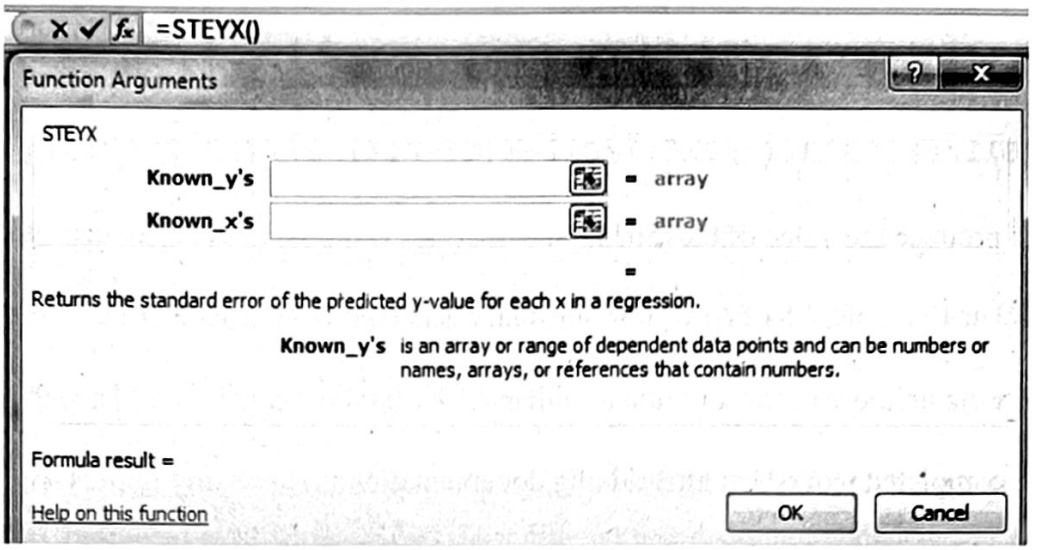

Figure 4-15 The STEYX Function Arguments window.

Select cells C2:C6 for the **Known_y's** and cells B2:B6 for the **Known_x's**. Click on OK. This should return the value of 0.144211 in cell B11. Note that this is the same value given by LINEST in cell F4 or by the Analysis ToolPak under Regression Statistics, Standard Error.

Now enter 5 for *N* in cell B12 or use the Count function as described earlier. In cell B13, we need to find the sum of the squares of the deviations of *x* from the mean *x*. This could be calculated from Equation 8-10 of FAC9 or 7-10 of AC7. The easiest way is to use Excel's built-in function **DEVSQ()** as discussed in Chapter 2. Thus, we enter in cell B13

=DEVSQ(B2:B6) [↵]

In cell B14, we need to find the average of the *y* values. To do this we type

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

```
=AVERAGE(C2:C6) [↵]
```

In cell B15, we type 1 or use the Count function for the number of repetitions of the unknown results. Now, we are only left with finding the standard deviation in c. We use Equation 8-18 of FAC8 or 7-18 of AC7, which is

$$s_c = \frac{s_r}{m}\sqrt{\frac{1}{M}+\frac{1}{N}+\frac{(\bar{y}_c-\bar{y})^2}{m^2 S_{xx}}}$$

Enter into cell B16, the formula corresponding to this equation

```
=(B11/B8)*SQRT(1/B15+1/B12+((C7-B14)^2)/(B8^2)*B13) [↵]
```

This should produce the value of 0.075633. Note that this is the same as calculated in Example 8-5 of FAC9 or Example 7-8 of AC7. For our final results, noting the standard deviation in c, we would report the unknown concentration as either 1.14 ± 0.08 or perhaps 1.144 ± 0.076 mole %.

The completed worksheet after adding documentation is shown in Figure 4-16.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.

	A	B	C	D	E	F
1		*x*	*y*			
2		0.352	1.09		2.092507	0.256741
3		0.803	1.78		0.134749	0.158318
4		1.08	2.6		0.987712	0.144211
5		1.38	3.03		241.1465	3
6		1.75	4.01		5.015089	0.062391
7	**Unknown**		2.65			
8	**Slope**	2.092507				
9	**Intercept**	0.256741				
10	**Concentration of unknown**	1.143729				
11	**Standard error in *Y***	0.144211				
12	***N***	5				
13	S_{xx}	1.145368				
14	***y* bar**	2.502				
15	***M***	1				
16	**Standard deviation in *c***	0.075676				
17						
18	**Documentation**					
19	Cell E2=LINEST(C2:C6,B2:B6,TRUE,TRUE)					
20	Cell B8=SLOPE(C2:C6,B2:B6)					
21	Cell B9=INTERCEPT(C2:C6,B2:B6)					
22	Cell B10=(C7-B9)/B8					
23	Cell B11=STEYX(C2:C6,B2:B6)					
24	Cell B12=COUNT(B2:B6)					
25	Cell B13=DEVSQ(B2:B6)					
26	Cell B14=AVERAGE(C2:C6)					
27	Cell B15=COUNT(C7)					
28	Cell B16=B11/B8*SQRT(1/B15+1/B12+((C7-B14)^2)/((B8^2)*B13))					

Figure 4-16 Completed least-squares worksheet after adding documentation.

The Internal Standard Method

In the *internal standard method*, a known amount of a reference species is added to all the samples, standards and blanks. The response signal is then not the analyte signal itself, but the *ratio* of the analyte signal to the reference species signal. A calibration curve is prepared where the *y* axis is the ratio of responses and the *x* axis is the analyte concentration in the standards as usual. The internal standard method is often used to compensate for errors if they influence both the analyte and the reference signals to the same proportional extent.

As an example, we'll use Example 8-7 of FAC9, which describes the determination of sodium by flame spectrometry. Lithium was added as an internal standard. The worksheet shown in Figure 4-17 can be constructed by entering the data from the table in Example 8-7 of FAC9.

© 2014 Cengage Learning. All Rights Reserved. May not be scanned, copied or duplicated, or posted to a publicly accessible website, in whole or in part.